

Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme

Heinz-Martin Süß, Klaus Oberauer & Martin Kersting*

Freie Universität Berlin und Deutsches Institut für Fernstudien an der Universität Tübingen

Zusammenfassung: Thema des Beitrags ist der Zusammenhang von Intelligenz, Wissen und komplexem Problemlösen. Theoretische Annahmen und methodologische Voraussetzungen zur Überprüfung des Zusammenhangs von Intelligenz und Problemlösen werden diskutiert, Ergebnisse einer Untersuchung mit 214 Probanden werden vorgestellt. Eingesetzt wurden ein Test für das Berliner Intelligenzstrukturmodell (BIS), Tests für allgemeines und problemspezifisches Wissen und dreimal das Problemszenario «Schneiderwerkstatt» unter Intransparenzbedingung.

Mit einer eingehenden Aufgabenanalyse wird begründet, daß tradierte Gütemaße für die «Schneiderwerkstatt» in der vorliegenden Untersuchung keine validen Indikatoren für die Steuerungsleistungen sind. Die Entwicklung eines neuen Problemlösegütemaßes wird detailliert dargestellt. Reliabilitätschätzungen und Zusammenhangsergebnisse zwischen Intelligenz und Problemlöseleistungen für Teilgütemaße, tradierte Gütemaße und das neue Gütemaß, werden berichtet. Die Annahme eines Zusammenhangs zwischen der intellektuellen Fähigkeit «Verarbeitungskapazität» und den Problemlöseleistungen konnte mit dem neuen Gütemaß gestützt werden.

Intelligence and control performance on computer-simulated systems

Summary: Theoretical assumptions and methodological prerequisites for investigating the interrelation between intelligence and complex problem-solving are discussed. The results of an empirical study with 214 subjects are presented. Subjects worked three times with the computer simulated problem-scenario «tailorshop» under conditions of intransparency. A test for the Berlin Model of Intelligence Structure (BIS) was used to assess different components of intelligence, and two tests were employed for measuring the amount of subjects' domain-specific knowledge.

A detailed task analysis of the «tailorshop» is conducted to establish that in the present study traditional indicators of control performance for the «tailorshop» could not be considered valid indicators. A new indicator of control performance and the rationale for its construction are described. Reliability estimates and correlational results for intelligence and control performance assessed by traditional as well as the newly developed indicators of control performance are reported. The assumption that there is in fact a relation between the intellectual capacity «complex information processing» and control performance was supported when the new indicators were used.

Sonderdruck aus

Sprache
*Zeitschrift für Sprach- und Kognitions-
psychologie und ihre Grenzgebiete*
& Kognition

Verlag Hans Huber Bern Stuttgart Toronto

1. Einleitung

Die Veröffentlichung von «Lohhausen» (Dörner, Kreuzig, Reither & Stäudel, 1983) hat die Forschungslandschaft nachhaltig beeinflusst: Problemlösen im Umgang mit computersimulierten realitätsorientierten Systemen wurde zum zentralen Gegenstand insbesondere der westdeutschen Problemlöseforschung.

Einige ungeklärte Probleme, die als Voraussetzung für weiterreichende Forschung einer Lösung bedürfen, werden im folgenden diskutiert:

1. Die neue Problemlöseforschung zeichnete sich aus durch die Hoffnung, über die angenommene «ökologische Validität» der com-

putersimulierten Szenarien alltägliches Problemlöseverhalten im Labor erforschen zu können. Die ökologische Validität der Untersuchungen mit diesen Problemszenarien ist aber bisher nur behauptet, nicht nachgewiesen worden. Dazu müßte gezeigt werden, daß die kognitiven Anforderungen, die ein Simulationsszenario an die Versuchspersonen stellt, eine repräsentative Stichprobe der Anforderungen an Problemlöser in einer definierten Problemdomäne ist (vgl. Pawlik, 1976). So bleibt der Verdacht, daß die Problemlöseforschung nicht nur neue wissenschaftliche Probleme, sondern auch neue Probleme für die Probanden geschaffen hat – Probleme, die mit Alltagssituationen nur oberflächliche Gemeinsamkeiten haben. Seit einigen Jahren finden formale Systeme, bei

* Herrn Prof. Dr. D. Dörner und einem anonymen Gutachter danken wir für ihre Anregungen.

denen auf eine semantische Einkleidung entweder ganz verzichtet wird oder ihr nur periphere Bedeutung zukommt, Verbreitung (z. B. Funke, 1986a). Erreicht wird dadurch, daß die interne Validität der Experimente steigt. Gleichzeitig wird auf den ursprünglichen Anspruch, Aussagen über Problemlöseverhalten in Alltagssituationen machen zu können, zumindest implizit wieder verzichtet. Eine ursprüngliche Intention der Forschung geht dabei verloren.

2. Die Reliabilität und Validität von Problemlösegütemaßen und anderen Verhaltensindikatoren ist nur unzureichend geklärt. Bisher gibt es lediglich einige Versuche, die Reliabilität von Problemlösegütemaßen zu bestimmen. Strohschneider (1986) berichtete «lower bound»-Schätzungen der Reliabilität eines Gütemaßes für das System Moro um $r=0.50$. Hasselmann und Strauß (1988) referierten Retest-Korrelationen zwischen $r=0.50$ und $r=0.74$ für verschiedene Gütemaße bei der wiederholten Steuerung der «Schneiderwerkstatt», sie verwendeten allerdings identische Systemstartwerte für die verschiedenen Bearbeitungsdurchgänge. Bruhn, Hasselmann und Strauß (1990) ermittelten für ein Gütemaß bei demselben System Korrelationen bis zu $r=0.89$ zwischen der ersten und der zweiten Bearbeitungshälfte. Dabei handelte es sich jedoch nicht um unabhängige Messungen.

Gütemaße werden häufig recht willkürlich definiert und haben oft keinen klaren Bezug zu dem Ziel, das den Probanden vorgegeben wurde (wenn etwa Expertenratings von Kurvenverläufen herangezogen werden, z. B. Dörner et al., 1983). Zu einer unüberwindbaren Barriere wird die Bestimmung adäquater Problemlösegütemaße dann, wenn Ziele offen oder polytelisch formuliert werden und ungeklärt bleibt, welche Ziele welche Probanden wie lange und mit welcher Gewichtung verfolgen.

3. Die Bedeutung von intellektuellen Fähigkeiten, wie sie in Intelligenzstrukturmodellen konzipiert sind, für die Vorhersagbarkeit von Steuerungsleistungen war Gegenstand einer heftigen Debatte. Ausgelöst wurde die Debatte durch die Befunde von Dörner et al. (1983) und Putz-Osterloh (1981), daß zwischen Intelligenztestleistungen und Leistungen bei der

Steuerung computersimulierter Systeme kein Zusammenhang bestehe. Dies sei ein Beleg für die mangelnde Kriteriumsvalidität der Intelligenztests. Die Gegenseite führte ins Feld, daß zunächst geklärt werden müsse, ob die Problemlösegütemaße reliabel und wofür sie valide sind (vgl. Funke, 1983; Tent, 1984; Hussy, 1985; Jäger 1986).

4. Neben der «Testintelligenz» rückte später Wissen über simulierte Systeme bzw. simulierte Gegenstandsbereiche in den Vordergrund des Forschungsinteresses. Die Grundannahme, daß Wissen eine Voraussetzung zur erfolgreichen Steuerung eines Systems ist, regte Forschungen sowohl zum Erwerb von Wissen während der Steuerung als auch zur Diagnose von Vorwissen an. Die Ergebnisse zum Zusammenhang zwischen dem Umfang richtigen Wissens und der Steuerungsgüte sind jedoch inkonsistent. Allerdings werden die angewandten Methoden der Wissensdiagnostik selten reflektiert.

In einem Teil der Forschungsbeiträge nimmt das Konzept des Wissens exklusiv die Rolle kognitiver Leistungsvoraussetzungen auf der Seite der Probanden ein und tritt dadurch implizit an die Stelle des Intelligenzkonzepts, anstatt mit ihm in eine theoretische Beziehung gesetzt zu werden. Dazu tragen besonders Begriffe wie «heuristisches Wissen» (z. B. Schaub & Strohschneider, 1989), die eine problemübergreifende Handlungskompetenz beschreiben sollen und somit offenbar denselben intendierten Geltungsbereich haben wie Intelligenzkonstrukte.

5. Die Beteiligung nicht-kognitiver Faktoren beim Problemlöseprozeß und ihr Einfluß auf das Ergebnis der Problemlösung haben schon Dörner et al. (1983) thematisiert. Einige Konstrukte (z. B. «Selbstsicherheit», «Ängstlichkeit») wurden seitdem als Kandidaten für die Aufklärung von Varianz sowohl der Problemlösegüte als auch bestimmter Verhaltensparameter diskutiert. Die vorliegenden empirischen Ergebnisse unterstützen für keines dieser Konstrukte konsistent die Hypothese eines Zusammenhangs mit Problemlöseleistungen (Kreuzig, 1981; Hesse, 1982; Reichert, 1986).

6. Erhebliche Probleme für die Vergleichbarkeit vorliegender Ergebnisse resultieren aus dem Einsatz ständig neuer Szenarien und Pro-

grammversionen ebenso wie aus Stichprobenproblemen. Die unsystematische Vermehrung der Simulationen sowie Gelegenheitsstichproben, wechselnd homogene und heterogene Kleinstichproben verhindern die Vergleichbarkeit von Untersuchungsergebnissen. Kluwe, Misiak und Haider (1991) halten unter anderem deswegen Metaanalysen für nicht durchführbar.

2. Fragestellung

Mit der hier vorgestellten Untersuchung wollen wir einige dieser Probleme angehen. Folgende Fragestellungen werden untersucht:

1. Wie reliabel sind Problemlösegütemaße?
2. Wieviel Varianz der Problemlöseleistungen können Intelligenztestwerte aufklären?
3. Wieviel inkrementelle Varianz von Problemlösegüteindikatoren können Wissensleistungen neben Intelligenztestleistungen aufklären?

Die Bestimmung der Reliabilität von Problemlösegütemaßen ist Voraussetzung für die Überprüfung der Zusammenhangsannahmen, weil die Reliabilität der Kriterien die Höhe der empirisch feststellbaren Zusammenhänge zu möglichen Prädiktoren begrenzt. Daher ist die Klärung der ersten Fragestellung den beiden anderen vorgeordnet. Wir erwarten, daß die Retest-Reliabilitätsschätzungen von Problemlösegütemaßen geringer sein werden als diejenigen von Intelligenztestaufgaben: Die computersimulierten Szenarien lassen den Probanden für die Lösung im Vergleich zu Intelligenztestaufgaben mehr Freiheitsgrade. Aufgrund der Abhängigkeit jedes Systemzustandes von den vorangegangenen können sich zunächst geringe nicht-intendierte Varianzquellen wie Motivationsschwankungen, Ermüdung, Ablenkung durch die Struktur des Systems «aufschaukeln» und über die meist sehr langen Bearbeitungszeiten kumulieren. Die lange Zeitdauer einer Problembearbeitung bringt daher keine Reliabilitätssteigerung durch Testverlängerung. Jeder Simulationsdurchgang liefert lediglich «single act»-Kriterien (Fishbein & Ajzen, 1974), deren Vorhersagbarkeit wegen ihrer geringen Reliabilität begrenzt ist. Mehrere Simulationsdurchgänge sind erforderlich, um die psychome-

trische Qualität der Problemlösegütemaße durch Aggregation zu sichern. Die Problemlösegütemaße aus einzelnen Bearbeitungsdurchgängen sind, unter anderem aufgrund des Wissenserwerbs der Probanden während der Bearbeitung, im Sinne der Testtheorie nicht vollständig äquivalent (Kluwe, Schilde, Fischer & Oellerer, 1991). Die Korrelationen zwischen wiederholten Durchgängen werden daher den Anteil systematischer Varianz in den Problemlösegütemaßen eher unterschätzen und können nur als «lower bound»-Schätzungen der Reliabilität gelten. Dennoch kann bei korrelativen Untersuchungen der Vorhersagbarkeit von Problemlöseleistungen darauf nicht verzichtet werden. Die (kaum bestreitbare) Feststellung, daß Handlungssequenzen beim Problemlösen nicht identisch wiederholbar sind (Dörner, 1991), impliziert nicht, daß zwei aufeinanderfolgende Problemlösungen derselben Person nichts mehr gemeinsam haben.

3. Theoretische Annahmen und Hypothesen

Probleme lassen sich analytisch zerlegen in die formale Problemstruktur einerseits und ihre semantische Einkleidung andererseits. Dies ist bei computersimulierten realitätsorientierten Problemen unmittelbar evident, gilt aber auch für Probleme der klassischen Problemlöseforschung. Kotovsky, Hayes und Simon (1985) haben dies am Beispiel des «Tower of Hanoi» gezeigt und zugleich nachgewiesen, daß eine Veränderung der semantischen Einkleidung bei gleicher formaler Problemstruktur starke Effekte auf die Schwierigkeit des Problems haben kann.

Komplementär zu dieser Analyse der Probleme lassen sich die kognitiven Voraussetzungen zur Lösung von Problemen einer Person in zwei Bereiche zerlegen: in den Bereich hochgradig genereller intellektuellen Fähigkeiten einerseits und in den Bereich relativ gegenstandsspezifischen Wissens andererseits.

In der Tradition der Intelligenzforschung wurden ursprünglich Fähigkeiten und Wissen begrifflich wie empirisch nicht differenziert. «Ability» war mit «aptitude» und «achievement», also wissensbasierten Leistungen, eng verbunden. Erst später wurden Fähigkeits- und Wissenskomponenten begrifflich und empirisch unterschieden.

- Das «testing the limits»-Konzept (Baltes & Kliegl, 1986) kann als Versuch interpretiert werden, die unterschiedlichen Wissensvoraussetzungen von Probanden dadurch auszugleichen, daß vor der Leistungsmessung Gelegenheit zu intensiver Übung besteht. Die danach verbleibende Leistungsvarianz differenziert dann die Probanden nur noch nach ihren intellektuellen Fähigkeiten.
- Die «investment theory» von Cattell (1971) mit der Unterscheidung von «kristalliner» und «fluiden» Intelligenz integriert stark wissensabhängige Kompetenzen als eigenes Teilkonstrukt unter dem Dach des Intelligenzbegriffs.
- Intelligenz wird als «Lernfähigkeit» (Guthke, 1974) konzeptualisiert und kann damit als Fähigkeit zum Erwerb von Wissen interpretiert werden. Fähigkeiten stehen in dieser Konzeption als Determinanten kognitiver Leistungen nicht mehr neben Wissen, sondern stellen die Voraussetzungen des Erwerbs von Wissen dar.
- Das Berliner Intelligenzstrukturmodell - BIS (Jäger, 1982, 1984) - geht von der Annahme aus, daß jede Intelligenzleistung durch operative und inhaltsgebundene Komponenten bedingt ist. Die operative Modalität beschreibt und differenziert dabei eher den fluiden Teil der Intelligenz, die Inhaltsmodalität (mit verbaler, numerischer und figuraler Kompetenz) den auf hochgradig generelles Wissen reduzierten wissensbasierten Teil. Beide Komponenten sind als verschiedene Aspekte der Intelligenz konzeptualisiert und können nicht getrennt gemessen werden.

Wir betrachten im folgenden Intelligenz als Voraussetzung für den Erwerb und die Anwendung von Wissen. Die Fähigkeiten zu induktivem und deduktivem Denken, zur Erinnerung und zur Aufmerksamkeit sind notwendige Bedingungen des Erwerbs und der Anwendung von Wissen, die interindividuell variieren. Zwischen intellektuellen Fähigkeiten und spezifischem Wissen ist ein enger empirischer Zusammenhang dann zu erwarten, wenn äquivalente Lernbedingungen der Personen für den Erwerb dieses Wissens angenommen werden können.

Intelligenz und problemrelevantes Wissen lassen sich als kognitive Voraussetzungen für Problemlösen den beiden Aspekten eines Problems zuordnen. Wir nehmen an, daß es primär von der

formalen Struktur eines Problems abhängt, welche Komponenten der Intelligenz in welchem Maße zur Lösung eines Problems notwendig sind. Auf der anderen Seite hängt es im wesentlichen von der semantischen Einkleidung eines Problems ab, welches Vorwissen über welchen Gegenstandsbereich für die Problemlösung erfolgreich angewandt werden kann.

Die semantische Einkleidung aktualisiert bei den Probanden Vorwissen aus dem Bereich, aus dem die semantische Einkleidung gewählt ist. Ob richtiges Vorwissen über einen Realitätsbereich für die Systemsteuerung nützlich ist, hängt unter anderem davon ab, ob das System den Realitätsbereich korrekt simuliert.

Die Bearbeitung der formalen Struktur eines Problems, das in der Steuerung eines Variablen-netzes besteht, erfordert besonders zwei Formen von Denkprozessen: Sie erfordert induktives Denken, um aus der großen Menge verfügbarer Systeminformationen Regeln zu erkennen und Hypothesen über Eigenschaften von Variablen und Relationen abzuleiten. Und sie erfordert deduktives Denken, um diese Hypothesen zu überprüfen. Deduktives Denken ist auch erforderlich zur Anwendung von Wissen: Durch deduktives Denken lassen sich aus dem Wissen über Relationen zwischen den Variablen konkrete Handlungsabsichten für bestimmte Ziele ableiten.

Induktives und deduktives Denken sind kognitive Leistungen, die in der Intelligenzstrukturforschung von herausragender Bedeutung sind. «Induction» und «syllogistic reasoning» gehören nach French, Ekstrom und Price (1963) zu den bestreplizierten Faktoren der Intelligenz. Im BIS sind Aufgaben, die induktives und/oder deduktives Denken erfordern, der Operationsklasse «Verarbeitungskapazität» (K) zugeordnet.

Wir nehmen deshalb an, daß Intelligenztestleistungen Prädiktoren für Leistungen an computersimulierten Systemen sind. Wichtigster Einzelprädiktor sollte die Skala «Verarbeitungskapazität» des BIS sein. Gleichzeitig hängt der Steuerungserfolg bei computersimulierten Problemen mit einer semantischen Einkleidung auch vom Umfang des richtigen Vorwissens ab. Wir erwarten, daß Intelligenz und Wissen interkorrelieren, gleichwohl aber beide Prädiktoren zusammen substantiell mehr Varianz an den Kriterien erklären als jede der beiden Gruppen allein.

Außerdem erwarten wir, daß im Falle wiederholter Systemsteuerung das Vorwissen zur Aufklärung der Varianz der Steuerungsleistungen bei der ersten Problembearbeitung am meisten beiträgt. Bei späteren Bearbeitungen des gleichen Problems ist das ursprüngliche Vorwissen als Prädiktor von immer geringerer Bedeutung, weil die Probanden beim Umgang mit dem System neues systemspezifisches Wissen erwerben.

4. Untersuchungsdesign

Berichtet wird ein Ausschnitt der Ergebnisse einer Untersuchung mit Oberstufenschülern von West-Berliner Gymnasien. Sie betreffen Reliabilitätsschätzungen für die Problemlösegütemaße und den Zusammenhang von Intelligenz und Steuerungsleistungen sowie die multiple Vorhersage von Problemlöseleistungen durch Intelligenz und Wissen. Über Ergebnisse zur Wissensdiagnose und zum Wissenserwerb sowie zum Zusammenhang des Wissens mit den Problemlöseleistungen wird an anderer Stelle ausführlicher berichtet (vgl. Süß & Kersting, 1990; Süß, Kersting & Oberauer, in Druck).

Der Untersuchungsplan ist in Tabelle 1 wiedergegeben.

4.1 Instrumente

Als Problemlöseszenario setzten wir das System «Schneiderwerkstatt» ein, das uns in der Fassung 2.3 von J. Funke vorlag. Die «Schneiderwerkstatt» gehört zu den am häufigsten eingesetzten Systemen. Bislang konnte für die «Schneiderwerkstatt» unter Intransparenzbedingung zu meist kein Zusammenhang zwischen Intelligenz- und Steuerungsleistungen aufgezeigt werden (Putz-Osterloh, 1981; Putz-Osterloh & Luer, 1981; Hörmann & Thomas, 1989). Eine Ausnahme ist eine jüngst veröffentlichte Studie von Hussy (1991), die jedoch auf einer Stichprobe von nur 20 Pbn beruht. Gegenüber vielen anderen Systemen hat die «Schneiderwerkstatt» den Vorteil, eine plausible semantische Einkleidung zu besitzen. Einige Fehler, auf die Funke (1986b) aufmerksam gemacht hat, haben wir korrigiert. Diese Korrekturen betrafen hauptsächlich die semantische Einkleidung. Die formale Struktur

des Systems haben wir so wenig wie möglich modifiziert, um die Anforderungen, die sich aus den Eigenschaften dieser Struktur ergeben, nicht zu verändern. Mit der Wahl des Problemszenarios sollte größtmögliche Vergleichbarkeit mit anderen Untersuchungen zum Zusammenhang von Intelligenz und Problemlösen hergestellt werden.

Die Probanden bearbeiteten das System unter «Intransparenzbedingung», d. h. sie hatten keine Informationen über die Vernetzung der Variablen, und mehrere Variablen blieben verdeckt. Sie bekamen die eindeutige Zielvorgabe, das Gesamtvermögen zu maximieren.

Zur Erhebung der Fähigkeiten setzten wir einen neuen, verbesserten Test für das Berliner Intelligenzstrukturmodell (BIS) ein. Das allgemeine Vorwissen erfaßten wir mit dem altersnormierten Wirtschaftslehrestest der Deutschen Gesellschaft für Personalwesen (DGP, 1986), ergänzt durch einige Fragen aus dem Wirtschaftslehrestest von Krumm und Seidel (1970). Zur Erfassung des spezifischen Vorwissens entwickelten wir ein neuartiges Erhebungsinstrument für die «Schneiderwerkstatt». Über diese Wissenstests berichten wir an anderen Stellen (Süß & Kersting, 1990; Süß, Kersting & Oberauer, in Druck).

4.2 Untersuchungsabfolge

Der experimentelle Untersuchungsplan sah drei Gruppen vor. Alle drei Gruppen bearbeiteten am ersten Tag den Intelligenz- und den Wirtschaftslehrestest.

Am zweiten Tag steuerten die Probanden die «Schneiderwerkstatt» dreimal: zunächst für 12 Simulationsmonate mit den üblichen Startwerten (vgl. Funke, 1986b), anschließend zweimal für 6 Monate mit jeweils veränderten Startwerten. Die Variation der Startwerte diente der Produktion von drei parallelen Testversionen, mit denen die Reliabilität des Systems «Schneiderwerkstatt» als Meßinstrument geschätzt werden kann. Die Verkürzung der Bearbeitungszeit auf 6 Monate war aus Zeitgründen notwendig. Sie erschien uns für die zweite und dritte Systembearbeitung vertretbar, da die Probanden bereits die Erfahrung aus dem ersten Durchgang nutzen konnten und daher keine Zeit zur Exploration des Systems verlieren mußten. Die Detailanalysen von Hussy (1991) zeigen, daß sich bereits nach

| Gruppe 1 N= 94 | Gruppe 2 N= 78 | Gruppe 3 N= 42 (Kontrollgruppe) |
|--------------------------------------------|--------------------------------------------|--------------------------------------------|
| ***** | ** 1. Tag ** | ***** |
| BIS-2 | BIS-2 | BIS-2 |
| WLT | WLT | WLT |
| ***** | ** 2. Tag ** | ***** |
| SWS-Einführung + 2 Übungsmonate | SWS-Einführung + 2 Übungsmonate | SWS-Einführung + 2 Übungsmonate |
| WIS-D1 | WIS-D1 | |
| | SWS-E (7 Variablen) SWS-D (9 Variablen) | |
| SWS-A (12 Monate) | SWS-A (12 Monate) | SWS-A (12 Monate) |
| WIS-D2 | WIS-D2 | WIS-D2 |
| SWS-B (6 Monate) SWS-C (6 Monate) | SWS-B (6 Monate) SWS-C (6 Monate) | SWS-B (6 Monate) SWS-C (6 Monate) |
| SWS-D (9 Variablen) SWS-E (7 Variablen) | | SWS-D (9 Variablen) SWS-E (7 Variablen) |
| WIS-D3 Kurzform | WIS-D3 Kurzform | WIS-D3 komplett |

Verzeichnis der eingesetzten Instrumente:

- BIS-2: Test zum Berliner Intelligenzstrukturmodell, Version 2.
- WLT: Wirtschaftslehrest. (DGP, 1986 und Krumm & Seidel, 1970).
- SWS: Simulationsprogramm «Schneiderwerkstatt» (Version von J. Funke, leicht modifiziert).
A / B / C bezeichnen die verschiedenen Bearbeitungsdurchgänge mit je neuen Startwerten.
D / E bezeichnen reduzierte Versionen mit neun bzw. sieben Variablen.
- WIS: Wissenstest. «D1»-«D3» bezeichnet die Meßzeitpunkte.

Untersuchungsgruppen:

- Gruppe 1: reduzierte Versionen der «Schneiderwerkstatt» am Ende der Untersuchung.
- Gruppe 2: reduzierte Versionen der «Schneiderwerkstatt» am Anfang der Untersuchung.
- Gruppe 3: Kontrollgruppe: kein Wissenstest vor der ersten SWS-Bearbeitung.

Stichprobenbeschreibung:

214 Oberstufenschüler (98 w, 116 m) von 13 West-Berliner Gymnasien.
Alter: Mittelwert = 17,6 (Standardabweichung = ,74).
Erhebungszeitraum: Oktober bis Dezember 1989.

5 Monaten deutliche Unterschiede zwischen erfolgreichen und weniger erfolgreichen Problemlösern abzeichnen.

Zusätzlich bearbeiteten die Probanden zwei in ihrem Umfang reduzierte Versionen mit 9 bzw. 7 vernetzten Variablen. Um Effekte der Darbietungsfolge zu kontrollieren, bearbeitete die erste Experimentalgruppe die reduzierten Systeme am Ende, die zweite Experimentalgruppe vor dem kompletten System. Auf die Ergebnisdarstellung zur Variation der Zahl vernetzter Variablen werden hier verzichtet und die Ergebnisse der beiden Experimentalgruppen werden aus Darstellungsgründen zusammengefasst. Diese Zusammenfassung ist vertretbar, da sich die Steuerungsleistungen der beiden Gruppen für das Gesamtsystem in keinem Bearbeitungsdurchgang signifikant unterschieden haben.

Das problemspezifische Wissen wurde in den Experimentalgruppen dreimal erhoben: vor und nach der ersten Systembearbeitung sowie in einer Kurzform am Ende. Zur Kontrolle der Reaktivität der Erhebung spezifischen Vorwissens auf die Problemlöseleistungen und auf die Wissenserhebung nach der Systembearbeitung arbeitete die Kontrollgruppe ohne den ersten Wissenstest. Über die Ergebnisse zu diesem Teil berichten Süß et al., 1990; Süß et al., in Druck.

Die Probanden steuerten die «Schneiderwerkstatt» nach zwei Übungsmonaten selbständig mit direkter Bedienung des Computers. Die Erhebung haben wir in Gruppen von 4 bis 8 Probanden durchgeführt, die jeweils unabhängig voneinander mit einem Computer arbeiteten.

4.3 Stichprobe

An der zweitägigen Untersuchung nahmen 249 Schülerinnen und Schüler der 12. Klassen von 13 Berliner Gymnasien teil, davon 214 an beiden Untersuchungstagen. Nur diese werden in der folgenden Darstellung berücksichtigt. Das Durchschnittsalter der 98 Schülerinnen und 116 Schüler betrug 17,6 Jahre.

5. Ergebnisse

5.1 Ergebnisse mit herkömmlichen Problemlösegütemaßen

5.1.1 Problemlöseleistungen

Um die Vergleichbarkeit mit vorliegenden Literaturergebnissen zu sichern, übernahmen wir zunächst zwei tradierte Problemlösegütekriterien: das Gesamtvermögen am Ende der Bearbeitungszeit (Hörmann & Thomas, 1989) und als Trendmaß die Zahl der Monate mit Gewinn (Funke, 1983). Tabelle 2 gibt einige deskriptive Ergebnisse für diese Gütekriterien wieder.

Im ersten Durchgang ist es keinem Probanden gelungen, das angestrebte Ziel zu erreichen, nämlich das Gesamtvermögen des Unternehmens zu steigern. Rund 44% der Probanden blieb unter dem Ergebnis des Null-Laufs des Systems. Das heißt, sie haben ein schlechteres Ergebnis erzielt, als wenn sie über alle simulierten Betriebsmonate überhaupt keinen Eingriff vorgenommen hätten. Nicht viel günstiger waren die Ergebnisse für das Trendmaß: Fast 70% haben in keinem Monat das Gesamtvermögen verbessert, gerade 2 von 214 Probanden in mehr als der Hälfte der Monate. Nur wenig besser waren die Ergebnisse im zweiten und dritten Durchgang.

Unsere Probanden, das belegen diese Ergebnisse sehr deutlich, waren mit dem System überfordert. Sowohl das Gesamtvermögen am Ende der Bearbeitung als auch das Trendmaß waren extrem schief verteilt. Daher haben wir die Variable Gesamtvermögen normalisiert und das Trendmaß dahingehend modifiziert, daß die Probanden für jeden Monat einen Punkt erhielten, in dem die Differenz des Gesamtvermögens zum Vormonat über dem Median der Gesamtstichprobe lag, berechnet über alle Eingriffsmomente. Diese beiden Gütemaße sind Grundlage der weiteren Ergebnisse (Tab. 3-5).

5.1.2 Reliabilität der Problemlösegütemaße

Die Reliabilität der Gütemaße (siehe Tab. 5), bestimmt als ihre Interkorrelation über die drei Bearbeitungsdurchgänge, war mit $r=0.37$ bis $r=0.53$ nicht besonders hoch und begrenzte damit die Vorhersagbarkeit deutlich. Dennoch kann man bei Reliabilitäten in dieser Größenord-

Tabelle 2: Steuerungsleistungen bei herkömmlichen Problemlösegütemaßen (Angaben in Prozent).

| | SWS-Bearbeitungsdurchgang | | |
|--------------------------------------------------|---------------------------|--------------------|--------------------|
| | SWS-A | SWS-B ¹ | SWS-C ¹ |
| Insgesamt Gewinn erwirtschaftet | 0,0 | 0,9 | 0,4 |
| Mehr als die Hälfte der Monate mit Gewinn | 0,8 | 3,4 | 5,1 |
| Kein Monat mit Gewinn | 69,8 | 73,2 | 65,5 |
| Besser als der Null-Lauf | 43,9 | 65,5 | 46,4 |

N=214;

Legende: siehe Tabelle 1

¹ Durchgang B und C: sechs Simulationsmonate

Tabelle 3: Korrelationen zwischen tradierten Problemlösegütemaßen und Wissen sowie den BIS-Skalen.

| | WLT | WIS-D1 | B | M | E | K | F | V | N | AI |
|----------------|------|--------|------|------|------|-----|------|------|------|------|
| GeKap-A | .03 | .02 | -.01 | .10 | -.06 | .13 | .05 | .07 | .01 | .05 |
| GeKap-B | -.02 | .04 | -.10 | .01 | -.04 | .01 | -.04 | .03 | -.09 | -.04 |
| GeKap-C | .11 | .07 | -.08 | .02 | -.05 | .08 | -.01 | -.01 | -.00 | -.01 |
| Trend-A | .09 | .09 | -.03 | .18* | -.02 | .15 | .12 | .11 | .02 | .10 |
| Trend-B | .00 | .11 | .03 | .08 | .02 | .07 | .01 | .08 | -.02 | .05 |
| Trend-C | .09 | .14 | -.04 | .07 | .03 | .10 | .08 | .01 | .04 | .05 |

N=214 (Spalte WIS-D1: N=149, nur Gruppen 1 und 2).

Legende: siehe Tabelle 1

Trend = Trendmaß mit Median als Trennwert

GeKap = Gesamtvermögen am Ende

B bis AI: Skalen zum Berliner Intelligenzstrukturmodell

- B = Bearbeitungsgeschwindigkeit
- M = Merkfähigkeit
- E = Einfallsreichtum
- K = Verarbeitungskapazität
- F = figürliches (anschauungsgebundenes) Denken
- V = verbales (sprachgebundenes) Denken
- N = numerisches (zahlgebundenes) Denken
- AI = Allgemeine Intelligenz (Integral aller Teilfähigkeiten)

nung auf systematische Varianz der Gütemaße schließen.

5.1.3 Zusammenhang zwischen Intelligenz- und Problemlöseleistungen

Tabelle 3 gibt die Korrelationen der 8 Skalen des Tests für das Berliner Intelligenzstrukturmodell

mit den Problemlösegütemaßen wieder. Die Ergebnisse scheinen eindeutig: Es gibt keine Zusammenhänge! Das gilt nicht nur für die Intelligenzleistungen, sondern auch für den Zusammenhang der herkömmlichen Gütemaße mit dem allgemeinen und dem spezifischen Vorwissen (Spalten 1 und 2).

Tabelle 4: Korrelationen der Teilgütekriterien und der Problemlösegütemaße.

| | Verk-A | Gekap-A | Trend-A |
|-----------------|---------|----------|----------|
| PLG-A | | .23* | .30** |
| GewSp-A | .51** | .61** | .62** |
| Verk-A | | -.20 | -.10 |
| | Verk-BC | Gekap-BC | Trend-BC |
| PLG-BC | | .23* | .34** |
| GewSp-BC | .26* | .74** | .72** |
| Verk-BC | | -.33** | -.15 |

N=127;

* p < .05 ** p < .01;

PLG = neues Problemlösegütemaß:

Summe aus Gewinnspanne pro verkauftem Hemd + Anzahl verkaufter Hemden

GewSp = Gewinnspanne pro verkauftem Hemd

Verk = Anzahl verkaufter Hemden

Gekap = Gesamtkapital nach dem letzten Monat

Trend = Trendmaß

A, B und C bezeichnen den Bearbeitungsdurchgang; BC = Aggregat aus Durchgang B und C

5.2 Ergebnisse mit einem neuen Problemlösegütemaß

5.2.1 Entwicklung eines neuen Problemlösegütemaßes

Dieses Ergebnis ist erklärungsbedürftig. Die gängige Erklärung für Nullkorrelationen zwischen Problemlöseleistungen und Intelligenztestwerten verweist auf eine Vielzahl von Unterschieden bezüglich der Anforderungen der beiden Aufgabenformen (z. B. Putz-Osterloh, 1981). Eine solche Erklärung ist unbefriedigend, solange nicht theoretisch begründet wird, warum die Merkmale, hinsichtlich derer sich Intelligenztest- und Problemlöseaufgaben unterscheiden, psychologisch relevant sein sollten.

Bevor aber eine theoretische Erklärung für diese Befunde gesucht wird, sollten zunächst trivialere Erklärungsmöglichkeiten geprüft und ausgeschlossen werden. Eine dieser Möglichkeiten ist, daß die Spezifikationen des Systems untauglich waren und dadurch die verwendeten Problemlösegütemaße - trotz der Nähe zur Zielvorgabe - nicht valide waren. Unter einem validen Gütemaß verstehen wir ein Kriterium, das das Ausmaß zweckrationalen problemlösenden Handelns der Probanden erfaßt.

Aus differenzierteren Analysen der Daten, die das Simulationsprogramm der «Schneiderwerkstatt» liefert, ergaben sich einige Hinweise:

Tabelle 5: Reliabilität des neuen und der tradierten Problemlösegütemaße sowie Korrelationen zwischen neuen und alten Gütemaßen.

| | PLG-B | PLG-C | GeKap-A | GeKap-B | GeKap-C | Trend-A | Trend-B | Trend-C |
|----------------|-------|-------|---------|---------|---------|---------|---------|---------|
| PLG-A | .61** | .52** | .26** | | | .30** | | |
| PLG-B | | .58** | | .36** | | | .36** | |
| PLG-C | | | | | .26** | | | .31** |
| GeKap-A | | | | .52** | .39** | | | |
| GeKap-B | | | | | .53** | | | |
| Trend-A | | | | | | | .47** | .37** |
| Trend-B | | | | | | | | .46** |

N=214

Legende: siehe Tabelle 4

Reliabilitätsschätzungen hervorgehoben

1. Alle Probanden machten insgesamt Verlust. Mehr als die Hälfte lag am Ende unter dem Ergebnis des Nullraufs des Systems.
2. Die Zahl der Eingriffe korrelierte negativ mit den Gütemaßen.
3. Indikatoren für die Größe des Betriebs (z. B. Anzahl der gekauften Maschinen, Anzahl der produzierten Hemden) korrelierten hoch negativ mit den untersuchten Gütemaßen: Vergrößerung des Betriebs schadet also!

Wir kamen dadurch zu folgender Annahme:
Bei den gegebenen Spezifikationen des Systems und unter den in der Problemsituation gegebenen Bedingungen führte vernünftiges Handeln der Probanden nicht zu Erfolg, sondern zu Mißerfolg.

Vernünftiges Handeln in der Problemsituation «Schneiderwerkstatt» ist es, zu versuchen, durch den Verkauf von Hemden Gewinne zu erwirtschaften. Die Gewinne sind das *Produkt* aus der Anzahl verkaufter Hemden und der Gewinnspanne pro verkauftem Hemd. Vernünftig ist es also, sowohl die Gewinnspanne pro Hemd zu steigern, als auch die Anzahl verkaufter Hemden. In diese zwei Teilziele läßt sich das Ziel, Gewinn zu machen, zerlegen.

Das Programm «Schneiderwerkstatt» ist so aufgebaut, daß Einnahmen nur durch den Verkauf von Hemden oder durch Zinsen erzielt werden können. Dabei sind die Zinsen eine vernachlässigbare Größe, so daß wesentliche Gewinne nur über den Verkauf von Hemden erwirtschaftet werden können.

Die «Gewinnspanne pro verkauftem Hemd» ist keine Variable, die im Programm «Schneiderwerkstatt» berechnet wird. Im Rahmen unserer Subziel-Analyse haben wir diese Variable für jeden Simulationsmonat wie folgt berechnet: die Veränderung des Gesamtvermögens im Vergleich zum Vormonat, dividiert durch die Anzahl verkaufter Hemden in diesem Monat. Dieselben Werte erhält man, wenn für jeden Monat von den Einnahmen die Summe aller Kosten subtrahiert und durch die Anzahl verkaufter Hemden geteilt werden.

Wenn nun die *Gewinnspanne pro verkauftem Hemd negativ ist*, erwirtschaften die Probanden mit jedem verkauften Hemd nicht Gewinn, sondern Verlust. Das bedeutet: je mehr sie verkaufen, desto mehr Verluste haben sie.

Genau das war bei unseren Probanden der Fall. Alle haben, gemittelt über alle Simulationsmonate, negative Gewinnspannen pro verkauftem Hemd.

Der Grund dafür, daß das Problem für unsere Probanden so schwierig war, liegt nicht darin, daß das System prinzipiell nicht zu steuern ist. Im Zustandsraum des Systems «Schneiderwerkstatt», also im n-dimensionalen Raum seiner möglichen Zustände (mit $n = \text{Anzahl der Variablen}$) gibt es aber nur einen schmalen Sektor, in dem die Gewinnspanne pro ver-

kauftem Hemd positiv ist. In diesen Sektor konnten unsere Probanden unter den von uns vorgegebenen Bedingungen das System offensichtlich nicht gezielt bringen. Es fehlte ihnen das spezifische Wissen, das für eine erfolgreiche Steuerung nötig gewesen wäre. Wesentliche Information, die ein «echter» Manager selbstverständlich hätte, beispielsweise über die Kosten einer neuen Maschine oder darüber, wieviele Arbeiter zum Bedienen einer Maschine erforderlich sind, werden den Probanden in der üblichen «intransparenten» Präsentationsbedingung vorenthalten.

Solange sich nun das System im Bereich negativer Gewinnspannen bewegt, maximieren Probanden, die den Verkauf maximieren, dadurch nicht ihre Gewinne, sondern Verluste.

Diese Analyse unterstellt zunächst vereinfachend, daß die Gewinnspanne pro verkauftem Hemd unabhängig ist von der Anzahl verkaufter Hemden. Diese Annahme trifft für das System «Schneiderwerkstatt» größtenteils zu. Im System «Schneiderwerkstatt» kann der Betriebsleiter nicht, wie in realen Betrieben, einfach beschließen, produzierte Waren nicht zu verkaufen. Der Verkauf von Hemden in einem simulierten Monat erfolgt «automatisch» in Abhängigkeit von Angebot (= Zahl Hemden im Lager) und Nachfrage. Angebote und Nachfrage hängen systemimmanent ab von den Ausgaben des Probanden für Produktion, Werbung, etc. Die Variable «Anzahl verkaufter Hemden» im System «Schneiderwerkstatt» ist also eine (direkte oder indirekte) Funktion aller anderen Variablen (außer der Variable Gesamtvermögen). Sie ist – bei *gegebenen* Relationen zwischen den Variablen – ein Indikator für die absolute Größe des Betriebs, die der Proband als Betriebsleiter selbst bestimmt. Die «Gewinnspanne pro verkauftem Hemd» ist andererseits ein Indikator für die Güte der Abstimmung der Variablen untereinander, also ein Indikator für den Grad der Optimierung der Variablenverhältnisse. Jedes suboptimale Verhältnis zwischen Variablen produziert «tote Kosten», Ausgaben, die keinen Effekt haben, weil andere Ausgaben nicht auf sie abgestimmt sind. Wenn beispielsweise ein Proband zu wenige Arbeiter für den vorhandenen Maschinenpark einstellt, weil er nicht weiß, daß für jede Maschine gerade ein Arbeiter notwendig ist, stehen teure Maschinen unproduktiv im Betrieb herum.

Hier wird deutlich, warum Expansion sich schädlich auf das Ziel, das Gesamtvermögen zu maximieren, ausgewirkt hat: Expansion erhöht zwar möglicherweise die Verkaufszahlen und damit auch die Einnahmen, sie erhöht aber notwendigerweise in ungefähr gleichem Maße die Kosten. Mehr Hemden zu verkaufen ist nicht möglich, ohne zugleich die Kosten für Produktion, Werbung, Verkaufsstellen usw. zu erhöhen. Daher führt eine Verbesserung der Verkaufszahlen in der Regel nicht zu einer Verbesserung der Gewinnspanne pro verkauftem Hemd. Wer beispielsweise alle Ausgaben gleichzeitig verdoppelt, verdoppelt damit auch den Verkauf – aber die Abstimmung zwischen den Variablen wird nicht besser. Die Gewinnspanne pro verkauftem Hemd bleibt nahezu konstant. Wenn ein Proband aber den Verkauf durch allgemeine Expansion seines Betriebes erhöht, ohne dabei die Gewinnspanne wesentlich zu ändern, vermehrt er damit seine Gesamtverluste.

Denken wir uns nun einen idealtypischen guten Problemlöser, einen Probanden also, der auf der Grundlage der ihm gegebenen Prämissen der

Problemsituation vernünftig handelt. Diese Prämissen sind: das vorgegebene Ziel, das in die beiden genannten Subziele zerlegbar ist, und die stark begrenzten Systeminformationen, die die Probanden bekommen.

Ein guter Problemlöser wird beide Teilziele gut erreichen: Er wird viele Hemden verkaufen, und er wird eine relativ hohe Gewinnspanne pro Hemd erzielen. Eine relativ hohe Gewinnspanne war aber aufgrund der spezifischen Systembedingungen bei *unseren* Probanden immer noch ein *negativer* Wert und bedeutet also: relativ *geringe Verluste* pro Hemd.

Wenn nun ein guter Problemlöser den Verkauf *und* die Gewinnspanne steigert, ist sein Gesamtverlust das Produkt aus einer *hohen* Zahl verkaufter Hemden und einem relativ *geringen* Verlust pro verkauftem Hemd.

Ein schlechter Problemlöser dagegen, der *wenig* verkauft und *hohe* Verluste pro verkauftes Hemd hat, wird insgesamt etwa gleich viel Verlust machen. Sein Gesamtverlust ist das Produkt aus einer geringen Zahl verkaufter Hemden und einem großen Verlust pro verkauftem Hemd.

Solange die Gewinnspanne pro verkauftem Hemd negativ ist, also mit jedem verkauften Hemd Verlust verbunden ist, ist das zweite Teilziel, den Verkauf zu steigern, kontraproduktiv. Gute Problemlöser heben also mit ihrem Erfolg beim zweiten Teilziel den Erfolg beim ersten Teilziel wieder auf.

Dadurch kann ein Gütemaß, welches das Gesamtvermögen zur Grundlage hat, in unserer Untersuchung nicht zwischen guten und schlechten Problemlösern unserer Stichprobe unterscheiden, denn erfolgreiches vernünftiges Handeln führte zu genauso großen Verlusten wie nicht vernünftiges Handeln. Gute Problemlöser, die bei beiden Teilzielen erfolgreich waren, haben im Endergebnis genauso schlecht abgeschnitten wie schlechte Problemlöser, die beide Teilziele weit verfehlt haben.

Diese Annahmen lassen sich empirisch prüfen. Aus ihnen folgt, daß die Variable «Gewinnspanne pro verkauftem Hemd» hoch positiv, die Variable «Verkauf» negativ mit dem Gesamtvermögen korrelieren müßte. Genau das ist der Fall. Tabelle 4 gibt die Ergebnisse für die drei Durchgänge der Problembearbeitung wieder:

5.2.2 Definition eines neuen Problemlöse-gütemaßes

Wir haben deshalb für unsere Untersuchung ein neues Problemlöse-gütemaß definiert. Dazu haben wir das Oberziel, das Gesamtvermögen zu steigern, in die zwei Subziele zerlegt:

1. Zahl verkaufter Hemden steigern
2. Gewinnspanne pro verkauftes Hemd steigern.

Die Veränderung des Gesamtvermögens ergibt sich in der Schneiderwerkstatt in jedem Monat, wie erwähnt, aus dem Produkt der Werte dieser beiden Variablen.

Die Werte dieser beiden Variablen wurden als Gütekriterien für die beiden Subziele bestimmt und zu einem Gesamtgütwert für die «Schneiderwerkstatt» aggregiert. Zuvor wurden die beiden Subgütekriterien wegen ihrer extrem schiefen Verteilung normalisiert.

Wenn unsere Annahme richtig ist, dann können die Gütewerte für beide Subziele mit Prädiktoren wie Intelligenz und Vorwissen korrelieren. Die *Summe* der beiden Teilgütwerte müßte dann ebenfalls mit den Prädiktoren korrelieren. Das *Produkt* jedoch aus einem *positiven* Wert für die Anzahl verkaufter Hemden und einem *negativen* Wert für die Gewinnspanne kann mit keinem externen Prädiktor aus dem Bereich kognitiver Leistungsvoraussetzungen korrelieren.

5.2.3 Reliabilität und Interkorrelationen der Gütekriterien

Tabelle 5 zeigt, daß die Reliabilitätskoeffizienten des neuen Problemlöse-gütemaßes mit Korrelationen von $r = 0.52$ bis $r = 0.61$ deutlich über den alten Werten liegen, die zum Vergleich mit angegeben sind. Sie bleiben jedoch wegen der bereits diskutierten Faktoren hinter denen von Intelligenztest zurück. Die Tabelle zeigt weiter, daß mit Korrelationen von $r = 0.26$ bis $r = 0.36$ der Zusammenhang zwischen dem neuen und den alten Problemlöse-gütemaßen – erwartungsgemäß – ziemlich gering ist.

5.2.4 Zusammenhang zwischen Intelligenz- und Problemlöseleistungen

Mit den (normalisierten) Gütekriterien für die beiden Teilziele und dem neuen Gütemaß haben wir die Korrelationen zu den BIS-Skalen noch einmal berechnet. Das Ergebnis zeigt Tabelle 6.

| | B | M | E | K | F | V | N | AI |
|-----------|------|-----|------|-------|-----|------|-------|-------|
| PLG-A | .09 | .16 | .08 | .43** | .15 | .16 | .32** | .26** |
| PLG-B | -.01 | .04 | -.01 | .25** | .08 | .06 | .10 | .10 |
| PLG-C | -.06 | .03 | -.08 | .32** | .09 | -.03 | .13 | .08 |
| PLG-BC | -.04 | .04 | -.05 | .33** | .10 | .01 | .13 | .10 |
| PLG-ABC | .01 | .10 | .00 | .40** | .10 | .01 | .13 | .10 |
| Verk-ABC | .05 | .10 | .03 | .35** | .14 | .07 | .24* | .18* |
| GewSp-ABC | -.04 | .07 | -.04 | .32** | .07 | .07 | .12 | .11 |

N=214
 Legende: siehe Tabellen 3 und 4

Es zeigten sich nun substantielle Zusammenhänge zwischen dem neuen PLG und der BIS-Skala K, «Verarbeitungskapazität». Die Höhe der Korrelationen muß verglichen werden mit dem durchschnittlichen «part-whole»-korrigierten Trennschärfekoeffizienten der einzelnen Aufgaben der Skala BIS-K ($r=.46$). Denn die Problemlösegütemaße sind als «single act»-Maße einzelnen Testaufgaben äquivalent. Von den übrigen BIS-Skalen wiesen nur N (zahlegebundene Fähigkeiten) und das Aggregat der Einzel Fähigkeiten, die «Allgemeine Intelligenz», signifikante Zusammenhänge zum Gütekriterium der ersten Systembearbeitung auf. Alle übrigen Skalen korrelierten überhaupt nicht mit den Gütekriterien. Deutlich ist, daß die spezifischere Fähigkeit «Verarbeitungskapazität» höher und konsistenter mit den Steuerungsleistungen korreliert als die Allgemeine Intelligenz. Das entspricht unseren Erwartungen. Beide Teilzielmaße zeigen im Prinzip dasselbe Korrelationsmuster: Die Korrelationen mit BIS-K betragen .32 für die Gewinnspanne und .35 für den Verkauf.

5.2.5 Wissen und Problemlösen

Mit hierarchischen multiplen Regressionen konnte gezeigt werden, daß sowohl systemspezifisches als auch allgemeineres betriebswirtschaftliches Vorwissen zusätzlich zu den Intelligenzmaßen signifikante inkrementelle Varianz

der Problemlöseleistungen aufklären. Die multiplen Korrelationen betragen $R=0.60$ für PLG-A als Kriterium und $R=0.51$ für das Aggregat aus PLG-B und PLG-C als Kriterium. Die Ergebnisse hielten einer Kreuzvalidierung stand. Über die Ergebnisse zum Wissen im Kontext des Problemlösens wird an anderer Stelle ausführlicher berichtet (Süß & Kersting, 1990; Süß et al., in Druck).

Diese Ergebnisse zeigen, daß differenzierte Intelligenztestwerte wichtige, aber nicht die einzigen bedeutsamen Prädiktoren für Problemlöseleistungen sind. Die Bedeutung von domänenspezifischem Vorwissen bei der Bewältigung von Problemen mit einer realitätsorientierten semantischen Einkleidung begrenzt – zusätzlich zur relativ geringen Reliabilität der Problemlösegütemaße – die zu erwartende Höhe der Korrelationen zwischen Intelligenztestleistungen und Problemlöseerfolg.

6. Diskussion

6.1 Diskussion des neuen Problemlösegütemaßes

Der Entwicklung reliabler und valider Problemlösegütemaße kommt als Voraussetzung zur Prüfung von Zusammenhangshypothesen entscheidende Bedeutung zu. Die Untersuchung hat gezeigt, daß die Validität eines Problemlösegütemaßes nicht nur von dem eingesetzten System, sondern auch von den Rahmenbedingungen seiner Darbietung abhängt. Die skizzierte Aufgabenanalyse zeigte, daß sogar Indikatoren, die sich unmittelbar am vorgegebenen Ziel orientierten, unter den gegebenen Bedingungen nicht valide zwischen mehr oder weniger rationalem Problemlösehandelns unterscheiden konnten. Dieses Problem entsteht dann nicht mehr, wenn es einer nennenswerten Teilstichprobe gelingt, das System in den Bereich positiver Gewinnspanne zu steuern. In diesem Fall müßten die von uns definierten Teilgütekriterien, das neue Gütemaß und die tradierten Gütemaße hoch interkorrelieren. Es gäbe verschiedene Möglichkeiten, diesen Fall zu realisieren: Verlängerung der Bearbeitungszeit (Zahl der Simulationsmonate, Zahl der Bearbeitungsdurchgänge u. a.), Vereinfachung des Problems oder Vermittlung zusätzlichen Wissens an die Probanden. Den dritten Weg haben wir zur Überprüfung der hier vorgenommenen Analyse in einer Wiederholungsstudie beschrieben (Süß, Kersting & Oberauer, 1991). Nach dieser Modifikation der Bedingungen ließ sich erwartungsgemäß die Äquivalenz des neuen Gütemaßes mit den direkt vom Ziel abgeleiteten Indikatoren zeigen.

Damit liegt die Vermutung nahe, daß auch andere veröffentlichte Ergebnisse, in denen kein Zusammenhang zwischen Intelligenztestleistungen und Steuerungsleistungen aufgezeigt werden konnte, Artefakte unreliabler und/oder invalider Gütemaße sind.

Unabhängig von dieser Analyse muß für künftige Forschungsarbeiten gelten: Gütekriterien für Steuerungsleistungen sollten in der Regel mit den vorgegebenen Zielen übereinstimmen. Diese Übereinstimmung ist im allgemeinen notwendige und hinreichende Voraussetzung für die Validität der Gütekriterien. Man kann davon ausgehen, daß diese Voraussetzung erfüllt ist, wenn es zumindest einer substantiellen Teilstichprobe gelingt, ein System zielgerichtet erfolgreich zu steuern. Selbstverständlich hätte die Brauchbarkeit des Systems «Schneiderwerkstatt» für die untersuchte Fragestellung vorab geprüft werden können. Entsprechende Modifikationen des Systems und/oder der Untersuchungsbedingungen hätten danach vorgenommen werden können. Wir haben darauf verzichtet, um die Vergleichbarkeit unserer Ergebnisse mit den bereits vorlie-

genden Ergebnissen in der Literatur sicherzustellen. Nur über den hier beschrittenen Weg konnte auf die möglichen Reliabilitäts- und Validitätsprobleme anderer Untersuchungen zum Zusammenhang von Intelligenz und Steuerungsleistungen beim System «Schneiderwerkstatt» aufmerksam gemacht werden. Eine ursprünglich anvisierte Reanalyse publizierter Datensätze anderer Autoren war leider nicht möglich, da die Autoren uns aus verschiedenen Gründen ihre Daten nicht überlassen konnten.

Unter diesen spezifischen Bedingungen war die Einführung eines neuen Problemlösegütemaßes notwendig. Dem neuen Problemlösegütemaß kommt damit die Funktion eines «Notnagels» zu, auf den bereits in der Wiederholungsuntersuchung, nach Modifikationen wieder verzichtet werden konnte (siehe Süß et al., 1991).

6.2 Diskussion der empirischen Befunde

Dörner hat 1989 (S. 45) erneut die Ansicht vertreten, daß die Korrelationen zwischen Problemlöseleistungen und Intelligenztestergebnissen etwa $r=0.10$ betragen würden und daß dieser Zusammenhang «für jegliche Prognostik oder Diagnostik ohne Wert» sei. Nach den hier berichteten Ergebnissen kann dagegen festgehalten werden, daß differenzierte Intelligenztestleistungen auch unter Intransparenzbedingung substantiell zur Vorhersage des Problemlöseerfolgs bei einem computersimulierten System beitragen. Auch Strohschneider (1991) berichtet einen signifikanten Zusammenhang zwischen BIS-K und dem Erfolg bei der Bearbeitung eines intransparenten Systems (MORO) unter der Bedingung exakter Zielvorgaben.

Die Validierung von Tests für die «allgemeine Intelligenz» an Leistungen beim Arbeiten mit computersimulierten komplexen Realitätsausschnitten verletzt in zweierlei Hinsicht die Bedingung der Symmetrie zwischen Prädiktor und Kriterium (Wittmann, 1988):

- Auf der Seite der Intelligenz ist es notwendig, in spezifischere Fähigkeiten zu differenzieren. Die Prädiktionskraft der hochgradig generalen und heterogenen Skala «Allgemeine Intelligenz» ist gering. Sie ist dadurch begrenzt, daß zwischen dem Kriterium für die Leistung bei

einem einzelnen Problem und der «Allgemeinen Intelligenz» bezüglich der Generalität keine Symmetrie besteht.

- Auf der Seite des Problemlösens ist der Varianzanteil des bereichsspezifischen Vorwissens an der Varianz der Problemlösegütemaße zu berücksichtigen. Die Prädiktionskraft von Intelligenztestmaßen für Kriterien, die wesentlich durch das Vorwissen der Probanden mitbedingt sind, ist von vornherein begrenzt.

Meßinstrumente für die «allgemeine Intelligenz» enthalten daher einerseits Varianz, die zur Vorhersage der Leistungen beim Lösen komplexer Probleme nicht notwendig ist. Andererseits enthalten Indikatoren für den Problemlöseerfolg Varianzanteile, deren Vorhersage mit Intelligenzkonstrukten nicht intendiert ist. Die Differenzierung intellektueller Fähigkeiten und die Hinzunahme von Vorwissen als Prädiktor kann zur Annäherung an eine symmetrische Konstellation von Prädiktoren und Kriterien beitragen.

Literatur

- Baltes, P. B. & Kliegl, R. (1986). On the dynamics between growth and decline in the aging of intelligence and memory. In: K. Poeck, H. Freund & G. Gänshirt (Eds.), *Neurology* (pp. 1-17). Heidelberg: Springer.
- Bruhn, C., Hasselmann, D. & Strauß, B. G. (1990). Bearbeitungsmerkmale eines komplexen, computersimulierten Problems als Indikatoren für die Problemlöseleistung. In: D. Frey (Hrsg.), *Bericht über den 37. Kongreß der DGFP in Kiel* (Bd. 1, S. 15-16). Göttingen: Hogrefe.
- Cattell, R. B. (1971). *Abilities, their structure, growth and action*. Boston: Houghton Mifflin.
- Deutsche Gesellschaft für Personalwesen (DGP) (1986). *Differenzieller Kenntnistest (DKT). Subtest Wirtschaft*. Hannover: DGP.
- Dörner, D. (1989). *Die Logik des Mißlingens*. Reinbek: Rowohlt.
- Dörner, D. (1991). *Über die Philosophie der Verwendung von Mikrowelten oder «Computerszenarios» in der psychologischen Forschung* (Working paper No. 7 der Projektgruppe Kognitive Anthropologie der Max-Planck-Gesellschaft).
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Huber.
- Fishbein, M. & Ajzen, I. (1974). Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 81, 59-74.
- French, J. W., Ekstrom, R. B. & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors*. Princeton, N.J.: Educative Testing Service.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung, oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 29, 283-302.
- Funke, J. (1986a). *Ein Forschungsprogramm zur subjektiven Repräsentation dynamischer Kleinsysteme: Aufbau und*

- Anwendung von Wissen in Abhängigkeit von Person- und Systemmerkmalen* (Berichte aus dem Psychologischen Institut der Universität Bonn. Band 12, Heft 1). Bonn: Friedrich Wilhelms Universität.
- Funke, J. (1986b). *Komplexes Problemlösen*. Berlin: Springer.
- Guthke, J. (1974). *Zur Diagnostik der intellektuellen Lernfähigkeit*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Hasselmann, D. & Strauß, B. (1988). Reliabilität von Leistungen bei der Bearbeitung von komplexen Problemlöseaufgaben. In: W. Schönplflug (Hrsg.), *Bericht über den 36. Kongreß der DGFP in Berlin* (Bd. 1, S. 68). Göttingen: Hogrefe.
- Hesse, F. W. (1982). Effekte des semantischen Kontextes auf die Bearbeitung komplexer Probleme. *Zeitschrift für experimentelle und angewandte Psychologie*, 29, 62-91.
- Hörmann, H. & Thomas, M. (1989). Zum Zusammenhang zwischen Intelligenz und komplexem Problemlösen. *Sprache & Kognition*, 8, 23-31.
- Hussy, W. (1985). Komplexes Problemlösen - Eine Sackgasse? *Zeitschrift für experimentelle und angewandte Psychologie*, 32, 55-74.
- Hussy, W. (1991). Komplexes Problemlösen und Verarbeitungskapazität. *Diagnostica*, 37, 314-333.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-226.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21-35.
- Jäger, A. O. (1986). Validität von Intelligenztests. *Diagnostica*, 32, 272-289.
- Kluwe, R. H., Misiak, C. & Haider, H. (1991). The control of complex systems and performance in intelligence tests. In: H. Rowe (Ed.), *Intelligence: Reconceptualisation and Measurement*. Hillsdale: N.J.: Erlbaum.
- Kluwe, R. H., Schilde, A., Fischer, C. & Oellerer, N. (1991). Problemlöseleistungen beim Umgang mit komplexen Systemen und Intelligenz. *Diagnostica*, 37, 291-313.
- Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). Why are some problems hard? Evidence from tower of Hanoi. *Cognitive Psychology*, 17, 248-298.
- Kreuzig, H. W. (1981). Über den Zugang zu komplexem Problemlösen mittels prozeßorientierter kognitiver Persönlichkeitsmerkmale. *Zeitschrift für experimentelle und angewandte Psychologie*, 28, 294-308.
- Krumm, V. & Seidel, G. (1970). *Wirtschaftslehrest (BWL)*. Weinheim: Beltz.
- Pawlik, K. (1976). Ökologische Validität: Ein Beispiel aus der Kulturvergleichsforschung. In: G. Kaminski (Ed.), *Umweltpsychologie: Perspektiven, Probleme, Praxis* (S. 59-72). Stuttgart: Klett.
- Putz-Osterloh, W. (1981). Über die Beziehung zwischen Teintelligenz und Problemlöseerfolg. *Zeitschrift für Psychologie*, 189, 79-100.
- Putz-Osterloh, W. & Luer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 28, 309-334.
- Reichert, U. (1986). *Die Steuerung eines nichtlinearen Regelkreises durch Versuchspersonen* (Memorandum Nr. 42 des Lehrstuhl Psychologie II der Universität Bamberg). Bamberg: Universität Bamberg.
- Schaub, H. & Strohschneider, S. (1989). *Die Rolle heuristischen Wissens beim Umgang mit einem komplexen Problem oder: Können Manager wirklich besser managen?*

- (Memorandum Nr. 71 des Lehrstuhls Psychologie II der Universität Bamberg). Bamberg: Universität Bamberg.
- Strohschneider, S. (1986). Zur Stabilität und Validität von Handeln in komplexen Realitätsbereichen. *Sprache & Kognition*, 4, 42-48.
- Strohschneider, S. (1991). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme. *Diagnostica*, 37, 353-371.
- Süß, H.-M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. *Diagnostica*, 37, 334-352.
- Süß, H.-M., Kersting, M. & Oberauer, K. (in Druck). Zur Vorhersage von Steuerungsleistungen an computersimulierten Systemen durch Wissen und Intelligenz. *Zeitschrift für Differentielle und Diagnostische Psychologie*.

- Süß, H.-M. & Kersting, M. (1990). *Bedeutung und Erfassung von Wissen für das Lösen komplexer Probleme*. Vortrag auf dem 37. Kongreß der DGFP in Kiel.
- Tent, L. (1984). Intelligenz und Problemlösefähigkeit. Kommentar zu Dörner, D. & Kreuzig, H. W.: Problemlösefähigkeit und Intelligenz. *Psychologische Rundschau*, 35, 152-155.
- Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In: J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology*, 2nd. ed. (pp. 505-560). New York: Plenum.

Heinz-Martin Süß, Deutsches Institut für Fernstudien an der Universität Tübingen, Konrad-Adenauer-Straße 40-44, D-72072 Tübingen
Klaus Oberauer, Martin Kersting, Institut für Psychologie, Forschungsschwerpunkt «Intelligenz und Wissen», FU Berlin, Bundesallee 187, D-10717 Berlin